Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v12n1p69

Weighted Lasso subsampling for high dimensional regression

By Uraibi

Published: 26 April 2019

Electronic Journal of Applied Statistical Analysis
Vol. 12, Issue 01, April 2019, 69-84
DOI: 10.1285/i20705948v12n1p69

Weighted Lasso subsampling for high dimensional regression

Hassan S. Uraibi*

University and department
Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq
Tel.: +964-7822700717

Published: 26 April 2019

Lasso regression methods are widely used for a number of scientific applications. Many practitioners of statistics were not aware that a small change in the data would result in an unstable Lasso solution path. For instance, in the presence of outlying observations, Lasso perhaps leads to an increase in the percentage of the false selection rate of predictors. The discussions on determining an optimal shrinkage parameter of Lasso are still ongoing. Therefore, this paper proposed a robust algorithm to tackle the instability of Lasso in the presence of outliers. The new weight function is proposed to overcome the problem of outlying observations. The weighted observations are for a certain number of subsamples to control the false Lasso selection. The simulation study has been carried out and uses real data to assess the performance of our proposed algorithm. Consequently, the proposed method shows more efficiency than LAD-Lasso and weighted LAD-Lasso and more reliable results.

keywords: Robust Lasso, LAD-Lasso, WLAD-Lasso, Subsamples, Outliers.

1 Introduction

Efron et al. (2004) proposed Least Angle Regression (LAR) serves as a non-greedy version of a forward selection method and is then collected with a forward stagewise and Lasso

*Corresponding author: hassan.uraibi@qu.edu.iq

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

penalty (Tibshirani, 1996) to find a Lasso solution path. However, the new version of Lasso has the ability to rank the most important variables but does not need to be significant, see (Khan et al., 2007; Brink-Jensen and Thorn Ekstrøm, 2014). In the last ten years, respectable scientific papers argue that in some cases lasso gives multiple solutions. Tibshirani et al. (2012) pointed out when the rank of the covariates matrix is less than the number of covariates, Lasso does not have a unique minimum. This case can happen where some of the covariates are drawn from a discrete probability distribution. Moreover, as a result of the small change that may happen in the data, drastic changes occur in the lasso solution paths such as inclusion and some noise variables. Instability in the Lasso path can also occur when Lasso penalty randomly picks single correlated covariates each time it is run. Consequently, repeating Lasso on the same data more than once yields different results.

Lasso is reputed to be sensitive in the presence of a heavy tail, which can occur in high dimensional data as a result of heterogeneity problems. Heterogeneity problem occurs due to sampling data from different subpopulations. It leads to a heavy tail in the distribution shape of the model errors (Wu and Ma, 2014). Therefore, characterizing the approximate distribution of the Lasso estimator becomes more complex because the distribution of the heavy tail is unknown but differs from the distribution of the bulk of data. However, this small change certainly results in a drastic change in Lasso solution paths. Many research efforts have been a dedicated to propose the robust version of Lasso that has an ability to deal with outliers in both directions, the design matrix X and response variable Y. Wang et al. (2007) proposed Least Absolute Deviation (LAD) of Lasso, which denoted as (LAD-Lasso) by combining $\sum_{i=1}^{n} |y_i - x_i^{'}\beta|$ LAD loss function with $n\sum_{j=1}^{p} \lambda_j|\beta_j|$ penalty where λ_j is adaptive tuning parameter for different coefficients β_j (Zou, 2006). LAD-Lasso is resistant to the presence of outliers Y direction, but it is breakdown when the leverage points (outliers in X direction) are present. Arslan Arslan (2012) proposed combining weighted LAD,see Ellis and Morgenthaler (1992), Hubert and Rousseeuw (1997 J.), Giloni et al. (2006a), Giloni et al. (2006b)$\sum_{i=1}^{n} w_i|y_i - x_i^{'}\beta|$ with adaptive Lasso penalty in one algorithmic framework which denoted as WLAD-Lasso. The weights $w_i = min\{1, \frac{p}{\mathcal{M}(x_i)}\}$ are chosen by using robust measure of distance $\mathcal{M}(x_i)$ to assign a down weight to leverage points. One of the important shortcomings of WLAD-Lasso is that it is less efficient than adaptive lasso when the errors distribution(no outliers and nor heavy tail) due to LAD-Lasso is not adapted for small errors (Lacroix (2011)).Rosset and Zhu (2007) employed LARS Efron et al. (2004) as a piecewise linear solution path through huberizing Lasso and improving the efficiency of estimation even no outlier in the data. However, the penalty function of Lasso needs to estimate the shrinkage parameter which is computed by using cross validation.It is well known that the cross validation method arbitrary splits data into two sets. Due to this arbitrary split lasso does not guarantee any reproducible results and the solution path will be instable.

Meinshausen et al. Meinshausen et al. (2009) found the subsampling technique that proposed by Politis and Romano (1994) can make an asymptotically correct inference around Lasso coefficients. Stability paths of subsampling technique is constructed from the probability for each covariate to be selected when randomly resampling the subsampling procedure on the same data many times. It is well evident that Lasso statistical inference can be obtained from multiple random splits without losing the asymptotic control of inclusion noise variables. The multi split procedure Meinshausen et al. (2009) repeats the method of single split Wasserman and Roede (2009) many of times. Based on the simple random sampling without replacement, the random single split procedure gives each observation in the data set the same chance of choice to be in the subsample. In this case, multi split procedure cannot avoid appearing the outliers in the subsamples. Therefore, multi split procedure would yield invalid p-values that cannot use in practice. Suggestion robustifying Lasso is an uncompleted solution with random subsamples procedure. Due to the percentage of outliers that may appear in the subsample under consideration, perhaps exceeds the breakdown point of the estimator.

This paper proposes identifying the true non-zero coefficients based on adjusted p-values for robust subsamples Lasso regression (PRS-Lasso) algorithm, that is to overcome the instability of Lasso solution path by weighting data of X and Y variables. The weights have been derived from reweighted multivariate normal location and scatter matrix, see (Olive and Hawkins (2010), Uraibi et al. (2017a), Uraibi et al. (2015), Uraibi et al. (2017b)) to downweight the outliers in multivariate normal data. This procedure is combined with the algorithm of finding the correct p-values of high dimensional regression Meinshausen et al. (2009). Meinshausen et al. (2009) assigned the value then extracted the non-zero coefficients are based on the corrected p-values in which 1 is assigned to zero coefficients and 0 to non-zero ones.

The rest of this paper is organized in the following, Section 2, illustrates the proposed algorithm that is described into subsections 2.1, c-steps concentration procedure to get the robust weights to the response variable and predictors and the Section 2.2, presents the procedure to get p-values for subsamples Lasso regression. In Section 3, the simulation study has been done to assess the performance of our proposed method with two robust methods. Carseats data with two modified Carseats datasets have been considered in Section 4 to illustrate the efficiency of our proposed method. A brief summary of this research follows Section 5.

2 P-values for Robust Subsamples Lasso Regression Algorithm

The first step of multisplit procedure is to split data into two random subsamples, the dimensional reduction such Lasso can be performed with the first subsample and end up with a set of p-values of regression coefficients which can be computed using LS method for the second subsample. Repeat this procedure for the certain numbers of times to get a set of p-values at each time. Finally, combining all sets of p-values and producing what is called correct p-values in which only the significant variables remain in the final model. The supsamples procedure has been discussed a lot in the statistical literature, for more details see (Buhlmann et al., 2013, Zhang and Zhang, 2014, Lockhart et al., 2014, Van de Geer et al., 2014,Javanmard and Montanari, 2013, Brink-Jensen and Thorn

Ekstrøm, 2014 and Meinshausen, 2015).

2.1 C-steps concentration procedure for weighting Y and X

Consider the high dimensional linear regression equation,

$$Y = X\beta + e \tag{1}$$

where Y is an $(n\times 1)$ response vector, X is $(n\times p)$ fixed design matrix of independent variables, β is an $(p\times 1)$ regression parameters vector and e is an $(n\times 1)$ random errors vector with iid. from the following contamination distribution $F(e) = (1-\epsilon)N(0,\sigma^2) + (\epsilon)G$, where G is another distribution different from $N(0,\sigma^2)$. Assume that $(n\times\epsilon)$ leverage points are present in each independent variable. Weighting X and Y has been done before multisplit procedure. Actually, the weights has been derived from robust location and scatter matrix Olive and Hawkins (2010) concentration algorithm as follows,

1. Combining Y and X in one frame to create multivariate data matrix, say $Z = [Y, X]$.
2. Find $\hat{\mu}_{RF}$ and $\hat{\Sigma}_{RF}$, the location and scatter estimators of, reweighted fast consistence and high breakdown Olive and Hawkins (2010). Consider that the critical value of desired upper bound of Mahalanobis distance is For $j = 1, 2$ Do
   a) The robust Mahalanobis distance which denoted as $\mathcal{M}(z_i)$ should be computed first, $\mathcal{M}(z_i) = (z_i - \hat{\mu}_{RF})'\hat{\Sigma}^{-1}_{RF}(z_i - \hat{\mu}_{RF})$
   b) Let $S = \{\sum \mathcal{M}(z_i) : \mathcal{M}(z_i) \leq \chi^2_{(0.975,p)}\}$, $\delta = \frac{(0.975\times n)}{(2\times S)}$ and then the upper tail of chai-sqaure critical value can be computed through $q = min(\delta, 0.995)$
   c) The new scatter matrix which is denoted as $\hat{\Sigma}_{RM}$ can be obtained by weighting $\hat{\Sigma}_{RF}$, $\hat{\Sigma}_{RM_{(j)}} = \left[\frac{Med(\mathcal{M}(z_i))}{\chi^2_{(\delta,p)}}\right] \times \hat{\Sigma}_{RF}$
   d) Let $\hat{\Sigma}_{RF} = \hat{\Sigma}_{RM_{(j)}}$ to re-weighted $\hat{\Sigma}_{RM_{(j)}}$ in the second replication.
   e) Next
3. Finally, the weights can be obtained from, $w_i = min\left[1, \frac{\chi^2_{(q,p)}}{\mathcal{M}(z_i)}\right]$ and the $Y_w = Y\times w$ and $X_w = X \times w$

2.2 p-values for Subsamples Lasso Regression

Let B is the total number of random splitting times of original data such that $b = 1, \ldots, B$ which is the indexing randomly splitting the original data into two disjoint groups of equal size. The Multi-split Meinshausen et al. (2009) algorithm can use the following procedure:
Stage I:
For $b = 1, 2, ..., B$

1. Let the original data be denoted as $D = \begin{bmatrix} Y_{w_1} & X_{w_{11}} & \dots & X_{w_{p1}} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{w_n} & X_{w_{1n}} & \dots & X_{w_{pn}} \end{bmatrix}$ randomly split D into two disjoint groups of equal size $(n/2)$ denoted as $D_{in}^{(b)} = (Y_{w_{in}}, X_{w_{in}})$, $D_{out}^{(b)} = (Y_{w_{out}}, X_{w_{out}})$ respectively.

2. Let $\hat{S}_{\mathcal{H}}^{(b)} = \{j; \hat{\beta}_j^{\mathcal{H}} \neq 0\}$ be the $\hat{\beta}_j^{\mathcal{H}}(\lambda)$ estimates of Lasso based on $D_{in}^{(b)}$ such that $N = \hat{S}_{\mathcal{H}}^{c(b)} = \{j; \hat{\beta}_j^{\mathcal{H}} = 0\}$ the set of zero coefficients.

3. Fit the set of active predictors in $\hat{S}_{\mathcal{H}}^{c(b)}$ by using Leas squares (LS) based on $D_{out}^{(b)}$ subsample data and calculating corresponding p-values as follows,

$$\tilde{P_j}^{(b)} = \begin{cases} P_j & \text{if } j \in \hat{S}_{\mathcal{H}}^{(b)} \\ 1 & \text{if } j \notin \hat{S}_{\mathcal{H}}^{(b)} \end{cases} \tag{2}$$

and then without aggregated, adjusted $\tilde{P_j}^{(b)}$ values as

$$\hat{P}_j^{(b)} = min\left(\tilde{P_j}^{(b)} \times |\hat{S}_{\mathcal{H}}^{(b)}|, 1\right) \tag{3}$$

<u>Stage II:</u>
The stage I leads to B vectors of $\hat{P}_j$ values. To aggregate all of them $\hat{P}_j$ vectors, Meinshausen et al. (2009) suggested that the quantile γ can be written as follows,

$$Q_j(\gamma) = min\left[1, q_\gamma\left(\frac{\hat{P}_j^{(b)}}{\gamma}; b = 1, 2, ..., B\right)\right] \tag{4}$$

for any fixed $\gamma \in (0, 1)$ with lower bound at least equals to 0.05, where $q_\gamma(.)$ is the empirical quantile function. Selection the proper γ requires adding more correction to control the Family-wise Error (FWER) rate at level through the correction factor $1 - log(\gamma_{min})$ with upper bound 4. Consequentially, we can get the robust adjusted p-values from the following formula,

$$P_j^{rob} = min\left\{1, 1 - log(\gamma_{min}) \inf_{\gamma \in (\gamma_{min}, 1)} Q_j(\gamma)\right\} \tag{5}$$

From all values of P_j^{rob} just the predictors coefficients that possess, $P_j^{rob} \neq 1$, will be in the best model.

3 Simulation Study

Simulation scenarios have been done in this section to assess the performance of PRS-Lasso with LAD-Lasso and WLAD-Lasso method respectively. Five criteria are considered to compare the performances of PRS-Lasso, LAD-Lasso and WLAD-Lasso methods, over all 5000 generated dataset that are obtained by replications : (1) the average number of true zero coefficients (Zero.coef) , (2) the average number of non-zero true coefficients (N.Zero.coef), the (3) the False selection Rate (FSR), and (4) the criterion of Median of Mean Absolute Deviations (MMAD) residuals with (5) its standard deviation (SD). The mathematical formulation of MMAD can be defined as $\text{MMAD} = \text{Median}\left\{\frac{1}{n}\sum_{i=1}^{n}(|X\hat{\beta} - X\beta|)\right\}$. The best method is the one that selects the highest value of Zero.coef which is associated with lowest values of false selection variables FSR, MMAD and SD, respectively. The value of N.Zero.coef is related with the value of FSR in which the zero true coefficient is selected by certain method to be non-zero estimated coefficient, therefore the best method that possess the lowest FSR value have to select all true N.Zero.coef. However the FSR represents the false selection rate which need to be controlled.

All of PRS-Lasso, LAD-Lasso and WLAD-Lasso methods were run together in one simulation framework to find the p-values vectors of estimated coefficients without extract the non-zero coefficients. These vectors of are putted as rows in the $P_{M\times K}$ matrix, where M is the number of replications and K is the number of selection methods, hence,$M = 5000$ replication and $K = 3$ methods are considered to generate the random response variable y from linear regression model in the following equation,

$$y = X\beta + \sigma e \tag{6}$$

where X is a design matrix of p predictor which has been sampled from joint Gaussian marginal distribution with correlation structure $\rho = 0.5$. The distribution of random errors e is generated from the following contamination model,

$$F(e) = [(1-\epsilon)N + (\epsilon)G] \times \sigma \tag{7}$$

where ϵ is the contamination ratio, σ is a signal to noise which chooses to be 2, N and G stand for standard normal and heavy tail distributions, respectively. In this simulation the double exponential distribution which is so called $Laplace(0,1)$ and t- distribution with 2 degree of freedom $t(2)$ are considered from the family of heavt tail distributions. The ϵ of the observations of each predictors are randomly selected and then shifted by adding 10 value to create the leverage points, where $\epsilon = \{0.05, 0.10, 0.15, 0.20\}$. In order to verify the ability of PRS-Lasso method to control the selection rate, three simulation studies were carried out to simulate three sparse models based on the vector of true regression parameter β as follows,

Simulation1: $\beta = (\underbrace{2,2,2,2,2}_{5}, \underbrace{(0,\ldots,0)}_{20})$, $p = 25$, and $n = 50$.

Simulation 2: $\beta = (\underbrace{1.5,0,0,0,3}_{5}, \underbrace{0,2,0,0,4}_{5}, \underbrace{0,2,0,0,0}_{5}, \underbrace{(0,\ldots,0)}_{35})$, $p = 50$ and $n = 100$

Simulation 3: $\beta = (\underbrace{1.5,0,0,0,3,0}_{6}, \underbrace{0,0,2,0,0}_{5}, \underbrace{0,2,0,0,0,3}_{6}, \underbrace{(0,\ldots,0)}_{33})$,$p = 50$ and $n = 200$. Each simulation study simulates two sparse models relies the mixture distribution

Table 1: The FSR, average of Zero and non-zero coefficients MMAD, SD of Simulation 1.

Simulation 1		$G \sim Laplace(0,1)$			$G \sim t(2)$		
ϵ	Model	LAD-Lasso	WLAD-Lasso	PRS-Lasso	LAD-Lasso	WLAD-Lasso	PRS-Lasso
	Z.Coef	5.74	5.18	19.88	7.16	9.82	19.84
	NZ.Coef	19.26	19.82	5.12	17.84	15.18	5.16
0.05	FSR	0.570	0.590	0.005	0.510	0.413	0.006
	MMAD	1.56	1.38	1.07	1.51	1.29	1.11
	SD	0.28	0.23	0.10	0.16	0.12	0.11
	Z.Coef	5.88	6.38	19.92	6.04	7.20	19.94
	NZ.Coef	19.12	18.62	5.08	18.96	17.80	5.06
0.10	FSR	0.565	0.545	0.003	0.558	0.512	0.002
	MMAD	1.49	1.29	1.06	1.54	1.34	1.08
	SD	0.11	0.10	0.10	0.32	0.16	0.11
	Z.Coef	5.18	5.78	19.94	5.54	6.18	19.96
	NZ.Coef	19.82	19.22	5.06	19.46	18.82	5.04
0.15	FSR	0.593	0.569	0.002	0.578	0.50	0.002
	MMAD	1.53	1.35	1.11	1.59	1.41	1.07
	SD	0.13	0.10	0.10	0.62	0.63	0.10
	Z.Coef	5.28	5.56	19.96	5.16	5.68	19.96
	NZ.Coef	19.72	19.44	5.04	19.84	19.32	5.04
0.20	FSR	0.589	0.578	0.002	0.594	0.573	0.002
	MMAD	1.48	1.35	1.01	1.65	1.47	1.06
	SD	0.12	0.12	0.10	0.22	0.22	0.11

of $F(e)$ model, in which the distribution of outliers is $G \sim Laplace(0, 1)$ which is assigned to the first sparse model, and then $G \sim t(2)$ is added to the second one. Both models are examined with different ratios of outliers, $\epsilon = \{0.05, 0.10, 0.15, 0.20\}$ of n. The results of PRS-Lasso, LAD-Lasso and WLAD-Lasso methods of Simulations 1,2 and 3 are listed in Tables 1,2 and 3,respectively. The results have been shown that the performance of PRS-Lasso method to extract the true Z.coef and NZ.coef is the best than others. As can be seen in Fig 1 and 2 most NZ.coef of LAD-Lasso and WLAD-Lasso methods distributed around the horizontal dotted line which represents the Z.coef, while PRS-Lasso forced most of them to be zero. Moreover, It can be observed the values of FSR of PRS-Lasso method are getting between 0.002 and 0.006. This means from 2 to 6 times, a type I error (overfit) can occur among 1000 models selected by the PRS-Lasso method, while the false selection of LAD-Lasso and WLAD-Lasso methods exceeds 500 times. When $p = 25$ predictors are simulated in Simulation 1 in which only 5 of predictors are having non-zero true coefficients, LAD-Lasso and WLAD-Lasso methods are given more than 70% as a chance to zero true coefficients to be chosen as non-zero coefficient. This selection opportunity certainly results in to increase the overfitting problem across overall replications.

It can notable the gap between the values of Z.coef that are identified by PRS-Lasso method and others. It is very big as Table 1,2 and 3 reported. All simulation studies have not recorded any underfitting case and three methods are included all non-zero true coefficients in the best model selection. For the resulting plot, see Fig. 2 and 3, across overall 5000 replications of Simulation 1, LAD-Lasso, WLAD-Lass and PRS-Lasso are included all non-zero true coefficients, even though LAD-Lasso, WLAD-Lass are suffered from overfitting problem. However, controlling the overfitting problem is crucial to determine the best method, therefore, LAD-Lasso and WLAD-Lasso methods are considered unstable selection method. In other words, the PRS-Lasso method is strongest for controlling the False Selection Rate (FSR).

Table 1, Table 2 and Table 3 reports the values of MMAD and SD are very close which shows the stability in the general performance of the three methods. This has not changed even though the sample size and the contamination distribution of random error are changed too. Hence, describing LAD-Lasso and WLAD-Lasso methods as stable methods does not mean their efficiency in the selection, but they continue on the same performance. On the other hand, the values of MMAD and SD of PRS-Lasso method shows more homogenous and more stable than the other methods. However, the PRS-Lasso method performs very well and reliable.

4 Revised Carseats Data

Carseats dataset is obtained from James et al. (2013). The 400 observations are collected from sale locations as well as ten potential predictors (seven quantitative and three qualitative). Only, quantitative predictors (CompPrice, Income, Advertising, Population, Price, Age and Education) have been chosen to construct two modified datasets. It is well known that Carseats data is clean (no outliers and leverage points), therefore it

Table 2: The FSR, average of Zero and non-zero coefficients MMAD, SD of Simulation 2.

Simulation 2		$G \sim Laplace(0, 1)$			$G \sim t(2)$		
ϵ	Model	LAD-Lasso	WLAD-Lasso	PRS-Lasso	LAD-Lasso	WLAD-Lasso	PRS-Lasso
	Z.Coef	14.42	18.06	44.94	15.82	17.88	44.90
	NZ.Coef	35.58	31.04	5.06	34.18	32.12	5.09
0.05	FSR	30.58	26.04	0.06	29.18	27.12	0.09
	MMAD	1.50	1.31	1.11	1.53	1.34	1.13
	SD	0.08	0.08	0.07	0.14	0.12	0.08
	Z.Coef	12.38	12.34	44.94	12.28	13.78	44.92
	NZ.Coef	37.62	37.66	5.06	37.72	36.22	5.08
0.10	FSR	32.62	32.66	0.06	32.72	31.22	0.08
	MMAD	1.53	1.37	1.15	1.59	1.42	1.14
	SD	0.09	0.08	0.08	0.11	0.11	0.10
	Z.Coef	10.38	12.20	44.92	10.64	11.78	44.94
	NZ.Coef	39.62	37.80	5.08	39.36	38.22	5.06
0.15	FSR	34.62	32.80	0.08	34.36	33.22	0.06
	MMAD	1.54	1.36	1.10	1.59	1.43	1.12
	SD	0.10	0.09	0.07	0.11	0.09	0.08
	Z.Coef	10.72	11.46	44.88	9.70	11.78	44.88
	NZ.Coef	39.28	38.54	5.12	40.30	38.22	5.12
0.20	FSR	34.28	33.54	0.12	35.30	33.22	0.12
	MMAD	1.55	1.40	1.09	1.70	1.50	1.06
	SD	0.07	0.07	0.07	0.37	0.36	0.08

Table 3: The FSR, average of Zero and non-zero coefficients MMAD, SD of Simulation 3.

Simulation 3		$G \sim Laplace(0, 1)$			$G \sim t(2)$		
ϵ	Model	LAD-Lasso	WLAD-Lasso	PRS-Lasso	LAD-Lasso	WLAD-Lasso	PRS-Lasso
	Z.Coef	13.66	16.90	44.92	14.38	17.74	44.92
	NZ.Coef	36.34	33.10	5.08	35.62	32.26	5.02
0.05	FSR	31.34	28.10	0.08	30.62	27.26	0.02
	MMAD	1.48	1.32	1.11	1.52	1.34	1.10
	SD	0.08	0.07	0.06	0.12	0.08	0.06
	Z.Coef	11.76	12.28	44.84	10.86	12.18	44.86
	NZ.Coef	38.24	37.72	5.16	39.14	37.82	5.14
0.10	FSR	33.24	32.72	0.16	34.14	32.82	0.14
	MMAD	1.51	1.36	1.13	1.55	1.40	1.13
	SD	0.09	0.09	0.08	0.12	0.11	0.07
	Z.Coef	10.30	11.62	44.92	9.20	11.84	44.92
	NZ.Coef	39.70	38.38	5.08	40.80	38.16	5.08
0.15	FSR	34.70	33.38	0.08	35.80	33.16	0.08
	MMAD	1.52	1.35	1.10	1.62	1.45	1.11
	SD	0.08	0.08	0.06	0.12	0.10	0.08
	Z.Coef	10.68	10.20	44.96	10.22	10.22	44.90
	NZ.Coef	39.32	39.80	5.04	39.78	39.78	5.10
0.20	FSR	34.32	34.80	0.04	34.78	34.78	0.10
	MMAD	1.56	1.39	1.07	1.68	1.54	1.05
	SD	0.09	0.08	0.08	0.23	0.22	0.09

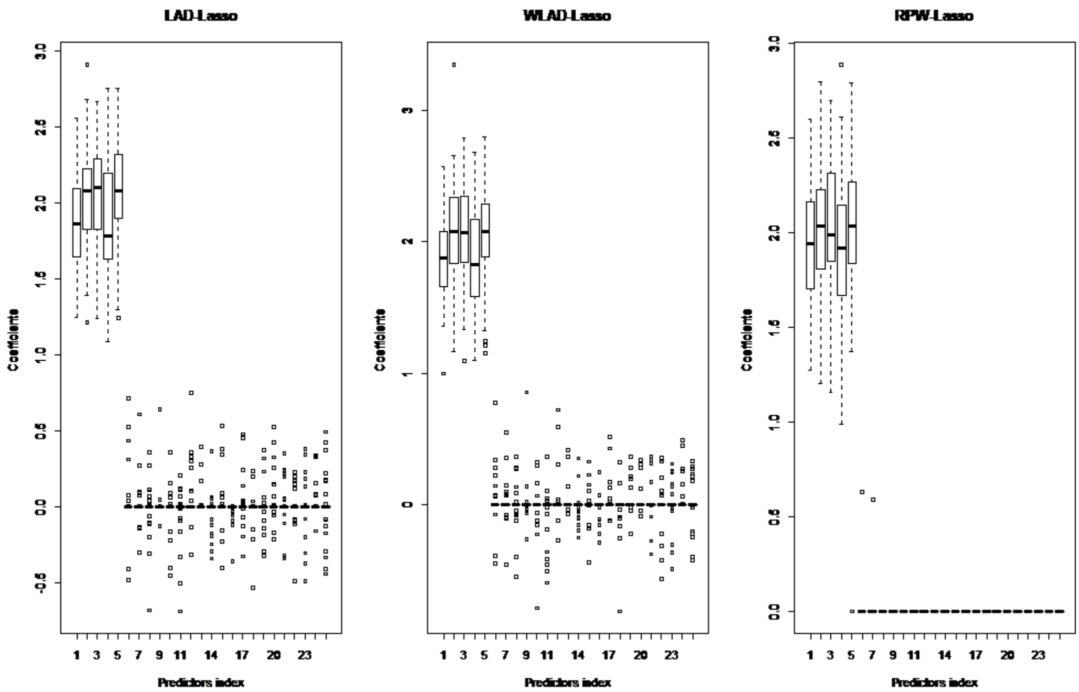

Figure 1: Boxplots of coefficients estimates from 5000 simulated datasets generated using simulation 1. Random errors contaminated by $Laplace(0, 1)$ distribution$n = 100$,$\epsilon = 0.2$, $\sigma = 2$. Horizontal dotted lines show the true zero values of the regression coefficients.

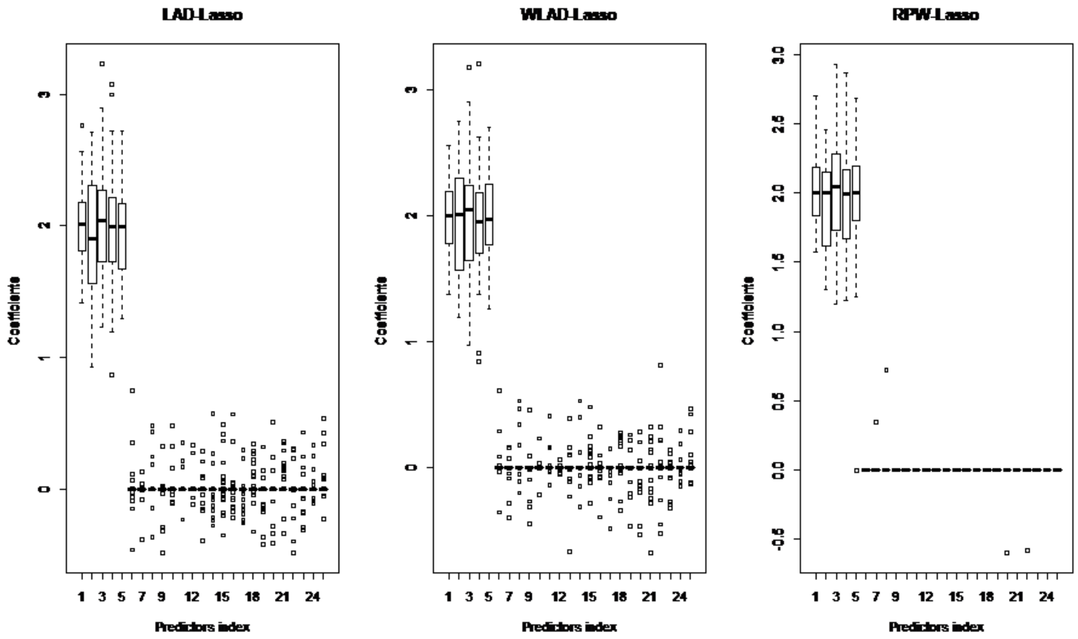

Figure 2: Boxplots of coefficients estimates from 5000 simulated datasets generated using simulation 1. Random errors contaminated by$t(2)$ distribution $n = 100$, $\epsilon = 0.2$,$\sigma = 2$. Horizontal dotted lines show the true zero values of the regression coefficients.

has been revised for two times. The first revision which denoted as (REV1) is to create 40 outliers in the Sales variable and 40 leverage points in population size predictor. The values of 40 clean observation in sales variable were enlarged through multiplying it by ten to create the outliers. Similar to this procedure, the predictor of population size has been contaminated to create 40 leverage points. The observations that were enlarged located at the top of the original data.

The second modified dataset which denoted as (REV2) involves increasing the dimensionality of data to be totally 30 predictor. The new 23 predictors were generated from normal standard distribution and the values of 40 observation at the top of the observations of each predictor have been enlarged ten times. The PRS-Lasso with LAD-Lasso and WLAD-Lasso methods are carried out using original and two revised datasets respectively. The picked up predictors by the certain method with original and REV1 datasets are configured two multiple regression models. The selected predictors of LAD-Lasso, WLAD-Lasso and PRS-Lasso are analyzed using LAD regression, WLAD-regression and Least Squares method respectively. The rq( ) from R-package quantreg is used to summarize the estimates of LAD regression, WLAD-regression methods. The R-code of PRS-Lasso method is combined some functions from hdi R-package with rmvn function that presents in David Olive website. The selected predictors of PRS-Lasso method are summarized using linear model lm( ) from MASS R-package. The analysis of predictors coefficients will reveal whether there is Type I or Type II error occurring in the selection of the predictors.

Table 4 is shown that the LAD-Lasso selects all predictors of Carseats dataset, but the summary of qr( ) where =0.5 reports the p-values of the coefficients of Population and Education predictors are not significant, therefore it is considered Overfit case. WLAD-Lasso identifies the coefficient of Education predictor is zero. The p-value of the Population predictor coefficient is non-significant in the summarized results of WLAD-Lasso Method. Table 4 presents also the selected predictors by PRS-Lasso method that diagnosed the true zero and non-zero coefficients of Carseats dataset predictors. Table 5 shows the results of three methods where 40 outliers and 40 leverage points are identified in sales variable and Population predictor of REV1 dataset, respectively. It is obvious that the LAD-Lasso identifies only the true zero coefficient of Education predictor, but the result of LAD method is summarized that the Advertising and Population are not significant too. The result WLAD-Lasso method has not been expected due to it fails to diagnose the true zero coefficients of Population and Education predictors, Overfit=2.

The PRS-Lasso performs very well and identified the true zero coefficients too without undefitting case. The existence of one or two zero predictors among the selected predictors by certain variable selection method may be acceptable. That due to, the statistical theories assumed that type I error may occur with variable selection method. However, the best variable selection method is that the one controlling Type I error. It is observed that the number of true non-zero coefficients in Carseats is five predictors and two are zero true coefficients, therefore it is not clear controlling the type I error at the lower bounds. The performances of three methods have been examined with REV2 dataset and the results are displayed in Table 6. The results thus obtained from table 6 are compatible with Its predecessors and confirmed that PRS-Lasso method outperformed all others. It is observed that PRS-Lasso method identifies the 5 true zero coefficients and 25 non-zero true coefficients among 30 predictors. On the other hand, LAD-Lasso overfits 11 predictors and identified only 4 true zero coefficients. The WLAD-Lasso may perform better than LAD-Lasso method, as shown in Table 6 in which the WLAD-Lasso identified 5 true non-zero coefficients and overfits 8 predictors without underfit problem.

5 Summary

This paper is a modest contribution to the ongoing improvements on instability Lasso solution path. Particular attention is paid to weight the X and Y as multivariate regression data. The weights are derived from multivariate normal location and scatter estimators to reduce the effect of outlying observation (outliers and leverage points). To control the false selection rate of Lasso and satisfying the stability, multisiplit procedure has been done for weighted data. Weighted data and multisplit procedure are combined in one computational algorithmic framework to proposed PRS-Lasso algorithm. The performance of proposed method is compared with LAD-Lasso and WLAD-Lasso.. The finding of the comparison was quite surprising and shows a high-efficiency in the PRS-Lasso, perfect controlling of a false selection occurring, a high stability solution path and further suggests it is hard to exclude true non-zero coefficients. Based on the findings of the current research, it is possible to conclude that our proposed method is more reliable than LAD-Lasso and WLAD-Lasso in the presence or not of outliers and leverage points in the data, so it has great potential and it can be readily used in practice.

References

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least angle regression. *The Annals of Statistics*, 32(2).

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, Series B, 267-288.

Khan, J. A., Van Aelst, S. and Zamar, R. H.(2007) Robust linear model selection based on least angle regression. *Journal of the American Statistical Association*, 102(480).

Wang, H., Li, G., Jiang, G. (2007) Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics* 25.

Brink-Jensen, K., Thorn Ekstrøm, C. (2014) Inference for feature selection using the Lasso with high-dimensional data. *eprint arXiv* 1403.4296.

Tibshirani, R. Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J., and Tibshirani, R J. (2012) Strong rules for discarding predictors in lasso-type problems *Journal of the Royal Statistical Society*, Series B, 74(2).

Wu, Cen and Ma, Shuangge (2014) A selective review of robust variable selection with applications in bioinformatics. *Briefings in bioinformatics*, 16(5).

Arslan, Olcay (2012) Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics* & *Data Analysis*,56(6).

Zou, Hui (2006) The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476).

Ellis, S P. and Morgenthaler, S. (1992) Leverage and breakdown in L 1 regression. *Journal of the American Statistical Association*, 87(417).

Hubert, M. and Rousseeuw, Peter J. (1997) Robust regression with both continuous and binary regressors. *Journal of Statistical Planning and Inference*, 57(1)..

Giloni, A., Simonoff, J. S. and Sengupta, B. (2006a) Robust weighted LAD regression. *Computational Statistics & Data Analysis*, 50(11).

Giloni, A., Sengupta, B. and Simonoff, J. S. (2006b) A mathematical programming approach for improving the robustness of least sum of absolute deviations regression. *Naval Research Logistics*, 53(4).

Lacroix, S.(2011) Robust regression through the Huber's criterion and adaptive lasso penalty. *Electronic Journal of Statistics*,5.

Rosset, S. and Zhu, J. (2007) Piecewise linear regularized solution paths. *The Annals of Statistics*, 1012–1030.

Meinshausen, N., Meier, L. and Bühlmann, P.(2009) P-values for high-dimensional regression. *Journal of the American Statistical Association*, (104).

Politis, D. N and Romano, J. P.(1994) Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*,2031–2050.

Wasserman, L. and Roeder, K.(2009) High dimensional variable selection. *Annals of Statistics*, 37(5A).

Olive, D. J. and Hawkins, D. M. (2010) Robust multivariate location and dispersion, *Preprint, see (www. math. siu. edu/olive/preprints.htm)*.

Uraibi, H. S., Midi, H. and Rana, S. (2017) Selective overview of forward selection in terms of robust correlations. *Communications in Statistics-Simulation and Computation*, 46(7).

Uraibi, H. S., Midi, H. and Rana, S. (2015) Robust stability best subset selection for autocorrelated data based on robust location and dispersion estimator. *Journal of Probability and Statistics*,2015.

Uraibi, H. S., Midi, H. and Rana, S.(2017) Robust multivariate least angle regression. *SCIENCEASIA*, 43(1).

Bühlmann, P. and others(2013) Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4).

Zhang, C. H. and Zhang, S. (2014) Confidence intervals for lowdimensional parameters in high-dimensional linear models. *Journal of the Royal Statistical Society*, Series B, 76.

Lockhart, R.,Taylor, J., Tibshirani, R. J. and Tibshirani, R.(2014) A significance test for the lasso. *Annals of Statistics*,42(2).

Van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. and others (2014) On asymptotically optimal confidence regions and tests for high-dimensional models em The Annals of Statistics,42(3).

Javanmard, A. and Montanari, A.(2013) Model selection for highdimensional regression under the generalized irrepresentability condition. *In Advances in Neural Information 2013. Processing Systems 26*.

Meinshausen, N. (2015) Group bound: confidence intervals for groups of variables in sparse high dimensional regression without assumptions on the design. *Journal of the Royal Statistical Society*, Series B,77(5).